A Picture is Worth a Thousand Search Results: Finding Child-Oriented Multimedia Results with *collAge*

Karl Gyllstrom
karl.gyllstrom@cs.kuleuven.be

Marie-Francine Moens
sien.moens@cs.kuleuven.be

Department of Computer Science
Katholieke Universiteit Leuven
Leuven, Belgium

ABSTRACT

We present a simple and effective approach to complement search results for children's web queries with child-oriented multimedia results, such as coloring pages and music sheets. Our approach determines appropriate media types for a query by searching Google's database of frequent queries for co-occurrences of a query's terms (e.g., "dinosaurs") with preselected multimedia terms (e.g., "coloring pages"). We show the effectiveness of this approach through an online user evaluation.

Categories and Subject Descriptors

H.5.0 [**Information Interfaces and Presentation**]: General

General Terms

Experimentation, Human Factors

Keywords

children, Google, query suggestion, Mechanical Turk

1. INTRODUCTION

Web search engines are not highly usable by children and young audiences [4], in part because results are presented to users in the form of a list of web page summaries, with a limited integration of mixed media (e.g., images). We believe that children's experiences with search engines would benefit from a greater emphasis on multimedia results, particularly of media that are not typically presented as first-class entities. For example, while Google search results commonly present images, there is no specialized presentation of interactive and learning media such as mazes, puzzles, tracing/coloring pages, music sheets, and games. We refer to these specialized media results as *TotBytes*, and we are building a search interface called *collAge* in which they can play a strong role in children's search experiences. Our search engine web page currently acts as a thin wrapper between the user and Google search, where, for the user's query, TotBytes are presented alongside fewer traditional web page search results. Figure 1 depicts a mockup of our system, showing a potential interface displaying TotBytes in response to a user's query about dinosaurs.

SIGIR'10, July 19–23, 2010, Geneva, Switzerland.
ACM 978-1-60558-896-4/10/07.

One challenge in integrating TotBytes into collAge is in determining appropriate types of media results to return for a particular query. For example, while a music sheet image would be an appropriate TotByte for a query about piano music, it would probably not be appropriate for a query about biology. To address this problem, our work leverages a "crowd-sourcing" approach in which frequently issued queries are used to determine whether or not a query and media type are coherent. For example, the query "dinosaur coloring pages" is issued quite frequently (e.g., 577,000 times in Google's Suggestion database [3]), while, presumably, the nonsensical query "calculus coloring pages" is not (at least, it does not appear in Google's database).

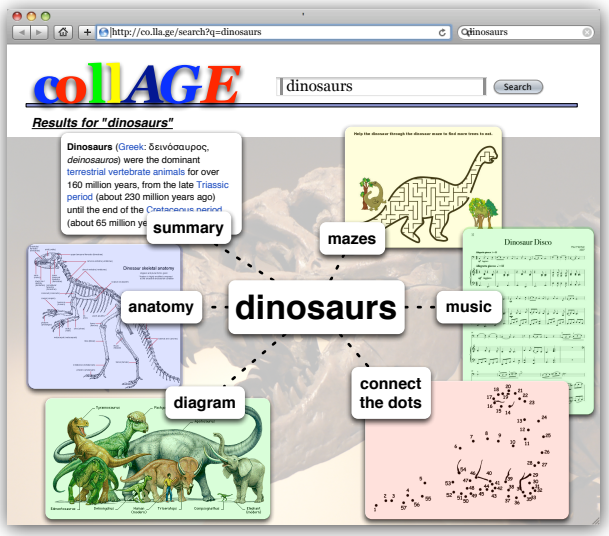

Figure 1: collAge mockup (can be zoomed).

2. APPROACH

Initially, we preselected a set of media types which we felt would be fun and useful for information-seeking children (see the left column of Table 1), although our approach can easily be expanded to new types. For a given user's query q, we generate a set of TotBytes TB_q as follows. First, for each media type, we create a media query $m_i \in M_q$ by appending to q the media type's terms (e.g., "crossword puzzles"). To filter out media types which are not meaningful complements to q, we dispatch each media query $m_i \in M_q$ – without its

last letter[1] – to Google Suggest [3], which generates a set of suggested queries $G(m_i)$. If $m_i \in G(m_i)$, it is considered to be a valid query (as m_i is frequently issued by Google users), and it is dispatched to Google's image search[2], from which the URL of the first result is added to the set of TotBytes for the user's query. The TotBytes can then be presented – with image links to the containing pages – to the user alongside traditional search results, as in Figure 1, although we leave the problems of presentation and ranking for future work. The combined processes of query suggestion lookup and image search take between 100-200 milliseconds on average, and the process for each media type can be executed in parallel.

3. EVALUATION AND RESULTS

We conducted an online evaluation with adults to determine if the TotBytes produced by collAge reflected the media type for which they were constructed; for example, that the TotBytes for the "music sheet" media type actually portrayed music sheets. Our hypothesis was that collAge would produce more accurate results than an approach that simply appended media terms without checking them against the database of popular queries. We first generated a set of queries Q_{eval} with which to evaluate our approach, as, to our knowledge, there is no existing corpus of children's web search queries. We constructed this set by collecting the titles of leaf subdirectories under the top-level topic "Kids and Teens" from the *Open Directory Project* [2], which included titles such as "dinosaurs" and "Egypt". Though not necessarily reflective of queries that children would naturally generate, these queries suffice in this stage of evaluation as we are merely determining the extent to which collAge can determine coherent TotBytes for arbitrary queries.

Next, we generated a set of TotBytes TB for a random sample of queries from Q_{eval}. For comparison, we also generated a set of baseline TotBytes, where the query database did not include their corresponding media queries, and added this set to TB. For each $b_i \in TB$, we constructed an HTML page displaying the media term(s) (e.g., "maps") and image result from b_i, a question asking if the image depicted an instance of the term (e.g., a map), and a "Yes/No" input form for users to record answers. These pages were uploaded to Amazon's Mechanical Turk [1] service, which presented the pages to human users, who answered their questions for a small payment. This produced a binary validity assessment for each b_i. We then compared the number of valid assessments between the collAge and baseline TotBytes across each media type, and measured the significance of their differences using the two-tailed Fisher's exact test.

Note that we assumed that a valid depiction of the media type partially validated that the media type was appropriate for the query. The theory was that less coherent combinations of queries and media terms are less likely to produce a result that depicts either query or media type. A limitation in this evaluation is that we do not ask for assessments for whether the TotBytes reflect the topic of the original query. This is due to the wide topical range of Q_{eval}; we felt users would often not recognize the query topic and use Google to learn about the topic, at which point they would be using the source of information (valid or not) to verify itself.

The results are depicted in Table 1. Our results included assessments by 156 unique Mechanical Turk worker IDs. Due to the anonymity policy, we could not collect demographic data on the users, but we assume from the site's policy that they are all at least 18 years of age. We filtered eligible candidates to those who have received favorable scores for at least 95% of their completed tasks.

Media type	collAge			baseline			P
	V	T	V/T	V	T	V/T	
music sheet	21	21	1.00	26	42	0.62	0.00
connect-the-dots	13	13	1.00	20	40	0.50	0.00
painting	103	112	0.92	25	41	0.61	0.00
coloring page	146	162	0.90	32	41	0.78	0.06
map	139	159	0.87	22	41	0.54	0.00
flag	82	98	0.84	16	41	0.39	0.00
anatomy	43	54	0.80	25	52	0.48	0.00
interactive game	16	21	0.76	26	46	0.57	0.17
maze	8	12	0.67	21	37	0.57	0.74
tracing page	2	3	0.67	29	51	0.57	1.00
word puzzle	9	15	0.60	15	36	0.42	0.36
crossword puzzle	7	23	0.30	14	49	0.29	1.00
Total	589	693	0.85	271	517	0.52	x

Table 1: Results. V indicates number of positive assessments, T indicates total assessments, and V/T indicates the ratio of valid to invalid assessments. P indicates P-value from the significance test between collAge and baseline.

4. CONCLUSIONS AND FUTURE WORK

collAge showed a consistent improvement over the baseline that was significant in cases where collAge was performing the strongest. The media types we chose had a range of effectiveness when used to generate TotBytes, with many performing quite well, and an overall strong performance. Even types at the low ranges could be useful (e.g., for a media type with 50% accuracy, simply displaying two TotBytes for the media type will create a 75% chance that at least one is meaningful.) Further, the collAge approach and evaluation can be easily repeated for new potential media types, making it generalizable to domains beyond children's interests. We plan to continue this research by implementing a usable prototype and evaluating it with children in a natural scenario, as well as determining further methods to find and rank media results.

Acknowledgements The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement No. 231507.

References

[1] Amazon Mechanical Turk. `http://www.mturk.com/`.

[2] ODP – Open Directory Project. `http://www.dmoz.org/`.

[3] Query Suggest FAQ. `http://labs.google.com/intl/en/suggestfaq.html`.

[4] D. Bilal and J. Kirby. Differences and similarities in information seeking: children and adults as web users. *Information Processing & Management*, 38(5):649 – 670, 2002.

[1]The last letter is removed from the query because Google Suggest will not offer a suggestion that is an exact match to the query provided.

[2]The results may be filtered for certain media types; for example, given a media query for the maze media type, we filter the result list to include only images that are grayscale and feature line art.